https://doi.org/10.2352/EI.2022.34.10.IPAS-354
© 2022, Society for Imaging Science and Technology

On properties of visual quality metrics in remote sensing applications

Oleg Ieremeiev[1], Vladimir Lukin[1], Krzysztof Okarma[2], Karen Egiazarian[3], Benoit Vozel[4]
[1] National Aerospace University, 61070, Kharkov, Ukraine;
[2] West Pomeranian University of Technology, 70-310, Szczecin, Poland
[3] Tampere University of Technology, FIN 33101, Tampere, Finland
[4] University of Rennes 1, 35042, Rennes, France

Abstract

Visual quality is important for remote sensing data presented as grayscale, color or pseudo-color images. Although several visual quality metrics (VQMs) have been used to characterize such data, only a limited analysis of their applicability in remote sensing applications has been done so far. In this paper, we study correlation factors for a wide set of VQMs for color images with distortion types typical for remote sensing. It is demonstrated that there are many metrics that have very high Spearman rank order correlation, e.g. PSNR-based and SSIM-based metrics. Meanwhile, there are also metrics that are practically uncorrelated with others. A detailed analysis of VQMs that have the largest SROCC values and belong to different groups is presented in this paper.

Keywords: visual quality, remote sensing image, correlation analysis, neural network

1. Introduction

Remote sensing (RS) imaging is widely used nowadays in numerous applications, such as ecology, forestry, agriculture, etc. [1, 2]. In many practical situations, acquired RS images are noisy and/or subject to other types of distortions (due to lossy compression, blur, etc.) [3-5]. These distortions can be inevitable due to the imaging system principle of operation (e.g., speckle in multichannel radar images [4, 6]) or the necessity to transform RS data in one or another way before visualization and final processing (e.g., to compress images before their transmission from satellites to on-land centers of RS data collection, processing and dissemination [7] or to carry out image registration and interpolation [8]).

RS images are often visualized and analyzed by experts [9. 10]. Therefore, image visual quality is of high importance, i.e. adequate VQMs should be used to characterize the visual quality of RS images.

Many VQMs of RS images have been proposed and analyzed in literature [11-15], both full-reference and no-reference ones. Meanwhile, there is no commonly accepted universal metric. This is due to many reasons. First, there are numerous types of RS images including single channel ones (similar to grayscale images), color or three-channel images and multi- and hyperspectral images, for which special ways of data visualization (e.g., in pseudo-colors) are employed. It is quite probable that different VQMs are optimal or quasi-optimal depending on applications. Second, depending on RS image type, different types of distortions are dominant. For example, speckle which a specific type of multiplicative noise-like phenomenon is the main factor which degrades quality of synthetic aperture radar (SAR) images [3, 6]. Spatially correlated noise can appear after geometric transformations used in image co-registration. Third, there is no established strict connection between the quality of RS images and the efficiency of solving further tasks of data processing, such as classification, object detection, etc. Fourth, there are no databases of distorted RS images that can be employed in the metric analysis and design.

If there is no commonly accepted visual quality metric, it is worth analyzing an applicability of the existing metrics and designing of new metrics [15]. Such an approach has been carried out in [15]. The database TID2013 [16] has been exploited – the color images with distortion types typical for remote sensing applications are extracted from this database.

Although such an approach has many limitations, it has allowed to find the best metrics according to their rank correlation with mean opinion scores (MOS). Note that none of the elementary metrics produces SROCC larger than 0.9 for all types of distortions present in TID2013. Due to this, one way is to use two or more elementary metrics in the analysis. However, it is worth using such metrics that have the correlation not too close to unity, i.e. metrics that differ from each other in their properties.

The combined metrics based on the trained neural network have been designed in [15]. In such a design, to keep the simplicity of a quality assessment, it is reasonable to apply a limited number of input elementary metrics. They must be efficient enough and, simultaneously, complementary with respect to each other, i.e. their correlation should not be high. Thus, one needs to have a priori information on cross-correlation properties of these metrics.

The main goal of this paper is to study properties of VQMs for RS applications.

Section 2 describes the results of preliminary analysis of VQM properties. Correlation analysis data for the considered VQMs are given in Section 3. The results of neural network VQM design are presented in Section 4.

2. Preliminary analysis

Let us recall the main results of the analysis carried out in [15]. The database TID2013 contains 25 test color images with 24 types and 5 levels of distortions. We were interested in two subsets of distortion types. The first one is called "Noise" and includes images with the following distortion types: Additive Gaussian noise (#1), Additive noise in color components (#2), Spatially correlated noise (#3), Masked noise (#4), High frequency noise (#5), Impulse noise (#6), Quantization noise (#7), Gaussian blur (#8), Image denoising (#9), Multiplicative Gaussian noise (#19), Lossy compression of noisy images (#21). Another subset, called "Actual", includes images with distortions ##1, 3, 4, 5, 6, 8, 9, 10 (JPEG compression), 11 (JPEG2000 compression), 19, and 21. In addition, it has also been proposed in [15] to consider jointly images for both subsets.

Table 1. SROCC values for 22 best metrics (according to the results for "Noise&Actual" subset)

Metric	All distortions	Noise	Actual	Noise & Actual
MDSI	0.8897	0.9275	0.9387	0.9374
CVSSI	0.8090	0.9248	0.9350	0.9341
MCSD	0.8045	0.9224	0.9326	0.9323
PSNRHA	0.8198	0.9230	0.9388	0.9322
GMSD	0.8004	0.9211	0.9314	0.9318
PSNRHMAm	0.8541	0.9221	0.9387	0.9315
PSIM	0.8926	0.9189	0.9309	0.9303
PSNRHAy	0.7794	0.9184	0.9272	0.9275
PSNRHVS	0.6536	0.9172	0.9257	0.9263
PSNRHMA	0.8137	0.9151	0.9343	0.9250
IGM	0.8023	0.9099	0.9220	0.9227
PSNRHMAy	0.7570	0.9107	0.9209	0.9226
VSI	0.8967	0.9101	0.9258	0.9218
SR-SIM	0.8076	0.9070	0.9211	0.9206
HaarPSI	0.8730	0.9063	0.9168	0.9190
ADM	0.7861	0.9113	0.9201	0.9189
PSNRHVSM	0.6246	0.9061	0.9175	0.9188
FSIMc	0.8510	0.9022	0.9150	0.9164
ADD_GSIM	0.8310	0.9023	0.9151	0.9159
IQM2	0.7955	0.8995	0.9103	0.9122
ADD_SSIM	0.8023	0.9008	0.9119	0.9120
FSIM	0.8011	0.8969	0.9108	0.9117

The SROCC values for all types of distortions and the considered subsets are presented in Table 1. The data are given for 22 elementary metrics that produce SROCC>0.9 for the subset "Noise&Actual" (in general, we have considered 50 elementary metrics, the references can be found in [17]).

Analysis of data given in Table 1 shows the following. First, several elementary metrics provide approximately the same efficiency of characterizing the visual quality of three-channel images with distortions types that take place in RS images. Second, these elementary metrics belong to different groups (families), such as PSNR-based metrics (as PSNRHA and PSNRHVSM) and SSIM-based metrics (as ADD SIM and FSIM). Meanwhile, there are also elementary metrics that do not belong to these families. This means that, on one hand, there are many efficient elementary metrics (although none of them is perfect) and, on the other hand, they are based on different principles and, thus, may complement each other. This ability has been, in fact, proven by the results of the combined metric design in [15]. It has been shown that a trained neural network that uses about 20 elementary metrics as inputs can provide SROCC for the subset "Noise&Actual" of about 0.965, i.e. sufficiently higher than SROCC of the best elementary metric.

Another problem with elementary metrics and image quality characterization is that the ranges of metric value variation are different. Some metrics are expressed in dB and vary in the wide limits, the others vary in the limits from 0 to 1. Most elementary metrics increase if the distortion level increases although there are also other cases. A linearization of MOS is often used to compare the metric's performance [15]. An example of such a linearization is given in Fig. 1. The scatterplot for each elementary metric shows pairs of metric value (after linearization) vs MOS for each considered test image, distortion type and level.

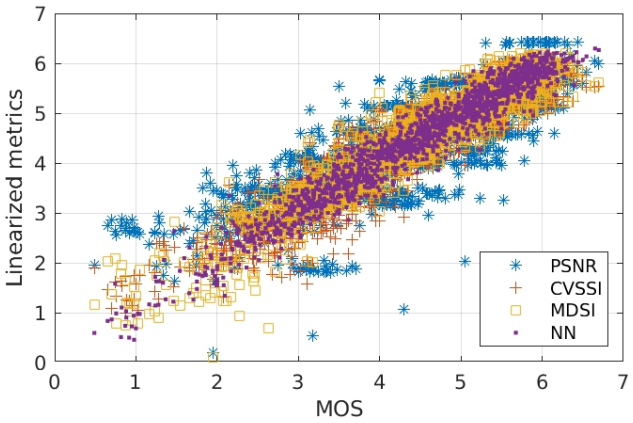

Figure 1. Scatter plot illustrating the correlation of four metrics (after linearization) vs true MOS (range 0-7) for TID2013 Noise&Actual subset.

Ideally, this should be a pure linear relation Linearized metric=MOS or, at least, a set of points placed closely and compactly to this line. However, as one can see, many PSNR points are placed far from this line. Some obvious outliers are observed even for the best elementary visual quality metrics MDSI and CVSSI. The NN-based combined metric provides higher compactness of the scatter-plot points. Thus, the preliminary analysis clearly indicated a necessity in correlation analysis of the elementary VQMs.

3. Correlation analysis results

We start our analysis from considering correlations between the conventional PSNR and other elementary metrics. Considering the non-linearity of all metrics,

SROCC values have been determined to minimize its influence. They are presented in Fig. 2. Data analysis shows the following. First, there are visual quality metrics which are highly correlated with PSNR (SROCC values exceed 0.8). These are mainly PSNR-based metrics (PSNRHVS, PSNR-HVSM, WSNR, and some others). Most other metrics (in particular, majority of SSIM-based ones) have rank correlation with PSNR about 0.7. This also relates to the top three metrics in Table 1 although SROCC for them is about -0.7 (this is because all three metrics reduce if the quality improves). Meanwhile, there are also other metrics such as WASH (that has SROCC about 0.3). Note that this metric has very poor performance for the subset Noise&Actual.

Next, we consider the correlation properties for metrics belonging to the same family. Table 2 presents data for a set of PSNR-based metrics. Data in Table 2 (as well as later in Tables 3 and 4) are put in the descending order of elementary metric SROCC with MOS.

Analysis of the results shows the following. First, for all PSNR-based metrics, their correlation with PSNR is high – the smallest SROCC is equal to 0.79 for the WSNR and VSNR. Second, SROCC values are also much higher for all pairs of visual quality metrics – all other pairs of metrics show greater results than mentioned ones. Third, there are pairs of metrics for which SROCC exceeds 0.9: WSNR and PSNRHVS, WSNR and PSNRHVSM, PSNRHVS and PSNRHVSM, PSNRHA and PSNRHMA, etc. This means that it is possible to analyze such metrics jointly but there is no expectation to benefit from their joint use. If one deals with a quality characterization for color or three-channel images, it seems reasonable to use PSNRHA that has high SROCC values for the considered subsets of distortions and large SROCC values for other PSNR-based metrics.

Table 3 shows data for the SSIM-based metric family. The analysis demonstrates that, first, the only metric for which its SROCC with SSIM is smaller than 0.7 is CWSSIM (0.67). Second, the only metric for which its SROCC with SSIM exceeds 0.9 is MSSIM. Other SSIM-based metrics have a high rank correlation with SSIM – mainly within 0.8 - 0.9.

One more conclusion is that all SSIM-based metrics are highly correlated. In particular, for the metric FSIMc (color version of FSIM), all SROCC values exceed 0.83, most SROCC values exceed 0.9 and the SROCC for FSIMc and FSIM is practically equal to unity (because of this, data for FSIM are not given in Table 3). Similar properties take place for the metric PSIM. Keeping in mind high values of SROCC for these metrics and MOS (see data in Table 1), it is reasonable to use either FSIMc or PSIM metrics from the considered family.

Thus, we have left PSNRHA, FSIMc, and PSIM for further analysis and added good metrics from Table 1 (those ones that produce SROCC with MOS larger than 0.9). The obtained SROCC values are given in Table 4. In analysis, we ignore signs of SROCC since they show only does a metric increase or decrease if image quality improves. Since all the metrics presented in Table 4 are "good", they have high values of SROCC between each other. For example, the absolute value of SROCC for MDSI and other metrics is not smaller than 0.963. SROCC values for PSNRHA and other metrics exceed 0.94.

Another observation is that there are pairs of elementary metrics for which SROCC approaches unity. This holds for CVSSI and MCSD, GMSD and MCSD. Taking this into account, we propose to leave for further analysis and joint use the following seven elementary metrics:

1) MDSI, since it provides the largest MOS for the considered types of distortions (see data in Table 1 or in Table 4);

2) PSNRHA, as the best PSNR-based metric that does not have too high correlations with other metrics;

3) GMSD, as a representation of the pair GMSD and MCSD having very high correlation;

4) PSIM, as one of the best SSIM-based metrics that does not have too high correlation with others in Table 4;

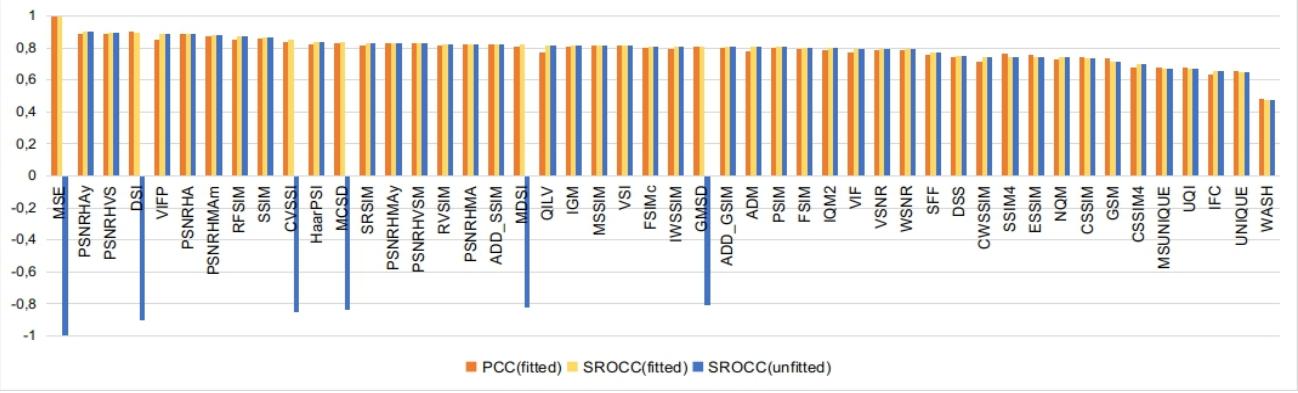

Figure 2. Metrics' SROCC with PSNR

Table 2. SROCC values for PSNR-based metrics

##	Metric	MOS (NA)	PSNR HA	PSNR HMAm	PSNR HVS	PSNR HMA	PSNR HVSM	WSNR	VSNR	PSNR	MSE
1	PSNRHA	0.93	1	1	0.97	0.99	0.96	0.93	0.91	0.89	-0.89
2	PSNRHMAm	0.93	1	1	0.97	0.99	0.96	0.93	0.91	0.88	-0.88
3	PSNRHVS	0.93	0.97	0.97	1	0.96	0.98	0.95	0.93	0.90	-0.90
4	PSNRHMA	0.93	0.99	0.99	0.96	1	0.97	0.94	0.91	0.83	-0.83
5	PSNRHVSM	0.92	0.96	0.96	0.98	0.97	1	0.97	0.93	0.83	-0.83
6	WSNR	0.90	0.93	0.93	0.95	0.94	0.97	1	0.90	0.79	-0.79
7	VSNR	0.88	0.91	0.91	0.93	0.91	0.93	0.90	1	0.79	-0.79
8	PSNR	0.83	0.89	0.88	0.90	0.83	0.83	0.79	0.79	1	-1
9	MSE	-0.83	-0.89	-0.88	-0.90	-0.83	-0.83	-0.79	-0.79	-1	1

Table 3. SROCC values for SSIM-based metrics

##	Metric	MOS (NA)	PSIM	SRSIM	FSIMc	ADD SSIM	IW SSIM	CS SIM	MS SIM	RF SIM	ES SIM	RV SIM	CW SSIM	SSIM
1	PSIM	0.93	1	0.98	0.98	0.96	0.95	0.97	0.94	0.93	0.94	0.91	0.84	0.80
2	SRSIM	0.92	0.98	1	0.99	0.97	0.97	0.97	0.96	0.95	0.93	0.95	0.84	0.85
3	FSIMc	0.92	0.98	0.99	1	0.97	0.98	0.98	0.97	0.93	0.94	0.95	0.83	0.85
4	ADD SSIM	0.91	0.96	0.97	0.97	1	0.98	0.96	0.97	0.91	0.91	0.91	0.81	0.87
5	IWSSIM	0.89	0.95	0.97	0.98	0.98	1	0.97	0.99	0.92	0.92	0.95	0.81	0.89
6	CSSIM	0.89	0.97	0.97	0.98	0.96	0.97	1	0.96	0.90	0.95	0.91	0.82	0.80
7	MSSIM	0.89	0.94	0.96	0.97	0.97	0.99	0.96	1	0.91	0.93	0.94	0.79	0.92
8	RFSIM	0.89	0.93	0.95	0.93	0.91	0.92	0.90	0.91	1	0.87	0.95	0.80	0.83
9	ESSIM	0.86	0.94	0.93	0.94	0.91	0.92	0.95	0.93	0.87	1	0.89	0.83	0.81
10	RVSIM	0.84	0.91	0.95	0.95	0.91	0.95	0.91	0.94	0.95	0.89	1	0.78	0.88
11	CWSSIM	0.81	0.84	0.84	0.83	0.81	0.81	0.82	0.79	0.80	0.83	0.78	1	0.67
12	SSIM	0.78	0.80	0.85	0.85	0.87	0.89	0.80	0.92	0.83	0.81	0.88	0.67	1

Table 4. SROCC values for several good metrics

##	Metric	MOS (NA)	MDSI	CVSSI	PSNR HA	GMSD	MCSD	PSIM	IGM	VSI	Haar PSI	FSIMc	ADD GSIM
1	MDSI	-0.937	1	0.976	-0.963	0.978	0.975	-0.985	-0.972	-0.988	-0.974	-0.985	-0.977
2	CVSSI	-0.934	0.976	1	-0.943	0.987	0.992	-0.974	-0.961	-0.96	-0.966	-0.972	-0.977
3	PSNRHA	0.932	-0.963	-0.943	1	-0.943	-0.945	0.959	0.951	0.966	0.965	0.952	0.941
4	GMSD	-0.932	0.978	0.987	-0.943	1	0.994	-0.983	-0.968	-0.963	-0.972	-0.975	-0.982
5	MCSD	-0.932	0.975	0.992	-0.945	0.994	1	-0.981	-0.966	-0.958	-0.973	-0.973	-0.985
6	PSIM	0.93	-0.985	-0.974	0.959	-0.983	-0.981	1	0.974	0.978	0.98	0.983	0.983
7	IGM	0.923	-0.972	-0.961	0.951	-0.968	-0.966	0.974	1	0.964	0.977	0.979	0.97
8	VSI	0.922	-0.988	-0.96	0.966	-0.963	-0.958	0.978	0.964	1	0.972	0.983	0.966
9	HaarPSI	0.919	-0.974	-0.966	0.965	-0.972	-0.973	0.98	0.977	0.972	1	0.986	0.974
10	FSIMc	0.916	-0.985	-0.972	0.952	-0.975	-0.973	0.983	0.979	0.983	0.986	1	0.983
11	ADD GSIM	0.916	-0.977	-0.977	0.941	-0.982	-0.985	0.983	0.97	0.966	0.974	0.983	1

5) VSI that does not have too high correlation with other metrics (Table 4);

6) HaarPSI that produces a quite high MOS (Table 1) but is not too highly correlated with others;

7) FSIMc that is one more good representation of SSIM-based metrics, but which is not too highly correlated with others.

4. Neural network based metrics

The use of neural networks (NN) is one of the possible ways to unite (combine, process jointly) a set of elementary metrics. Its performance depends on several factors: number of inputs, NN structure, neuron activation function, learning strategy, etc. One more factor is a preliminary processing of elementary metrics, particularly whether they are linearized or not? The results presented in [15] demonstrate the

influence of most of these factors. Our task here is not to carry out such a wide study. Instead, we would like to see the results of combining a fixed number of elementary metrics (in our case, seven) based on the performed analysis of their correlation properties.

To study the influence of the NN structure, we have analyzed six variants described in Table 5 in the column "Layers" (we give numbers of neurons in the hidden layers). Configuration of the layers of neural networks is determined based on the amount of available input data and the complexity of the task.

Data preparation was performed based on the results obtained in [15]. For example, fitting is not used for the selected set of elementary metrics, since the task of data normalization is effectively performed by choosing a sigmoid activation function.

A feed-forward network with two or three hidden layers was selected, as it is shown in Table 5. To ensure the maximum training accuracy, distorted images were dynamically distributed to the learning and training subsets in a ratio of 70% to 30%, 50 networks were trained for each configuration. The best results according to SROCC values obtained for them are shown in Table 5. The number of epochs is determined dynamically by a number of maximum validation failures during a decreasing backpropagation error.

Table 5. Data for the designed neural network based metrics with seven inputs

##	Layers	SROCC (train)	SROCC (test)
1	[7, 4]	0.9496	0.9514
2	[7, 4, 1]	0.9499	0.9516
3	[2, 2]	0.9483	0.9496
4	[4, 4]	0.9507	0.9494
5	[7, 7]	0.9460	0.9559
6	[7, 7, 7]	0.9493	0.9487

Analysis of data in Table 5 shows that there is a certain improvement of the combined metrics' performance compared to the best elementary metrics. There is no much difference between NNs with different structure, although variant 2 seems to be quantitatively the best. Meanwhile, the achieved SROCC values (about 0.95) are anyway slightly worse (smaller) than those reached in [15] for the larger numbers of elementary metrics (about 20) used as NN inputs. So, a trade-off between NN efficiency and simplicity can be found depending on the priority of requirements in practice.

Conclusions

An analysis of visual quality metrics with application to images with distortions typical for remote sensing applications is carried out. It is shown that there are many metrics that perform well. Correlation properties for these metrics have been analyzed using SROCC. It is shown that SROCC is high for many possible pairs of metrics.

This has allowed recommending the best metrics for PSNR and SSIM based families. It is also shown that the neural networks that use seven elementary metrics as inputs can produce SROCC with MOS about 0.95, i.e. at the very high level.

References

[1] S. Khorram, C.F. van der Wiele, F.H. Koch, S.A.C. Nelson, and M.D. Potts, "Future Trends in Remote Sensing," Principles of Applied Remote Sensing, Springer International Publishing: Cham, New York, 2016, pp. 277–285.

[2] N. Joshi, M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. Jepsen, T. Kuemmerle, P. Meyfroidt, E. Mitchard, and et al. "A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring," Remote Sensing, MDPI, vol. 8, no 1, pp. 1-23, 2016.

[3] C.-A. Deledalle, L. Denis, S. Tabti and F. Tupin, "MuLoG, or How to Apply Gaussian Denoisers to Multi-Channel SAR Speckle Reduction?," Transactions on Image Processing, vol. 26, no. 9, pp. 4389–4403, 2017.

[4] P. Zhong and R. Wang, "Multiple-Spectral-Band CRFs for Denoising Junk Bands of Hyperspectral Imagery," Transactions on Geoscience and Remote Sensing, vol. 51, no. 4, pp. 2260–2275, 2013.

[5] Y. Li and K.C. Clarke, "Image deblurring for satellite imagery using small-support-regularized deconvolution," Journal of Photogrammetry and Remote Sensing, vol. 85, pp. 148–155, 2013.

[6] S.G. Dellepiane and E. Angiati, "Quality Assessment of Despeckled SAR Images," in International Geoscience and Remote Sensing Symposium, Vancouver, pp. 3803-3806, 2011.

[7] I. Blanes, E. Magli and J. Serra-Sagrista, "A Tutorial on Image Compression for Optical Space Imaging Systems," Geoscience and remote sensing magazine, vol. 2, no. 3, pp. 8–26, 2014.

[8] T. Xu, and Y. Fang, "Remote sensing image interpolation via the nonsubsampled contourlet transform," in International Conference on Image Analysis and Signal Processing, Zhejiang, China, pp. 695–698, 2010.

[9] R. Sivakumar, "Image Interpretation of Remote Sensing data," Geospatial World, 2010, https://www.geospatialworld.net/article/image-interpretation-of-remote-sensing-data/

[10] "Visual Image Interpretation," https://fis.uni-bonn.de/en/recherchetools/infobox/professionals/image-analysis/visual-image-interpretation

[11] O.A. Agudelo-Medina, H.D. Benitez-Restrepo, G. Vivone and A. Bovik, "Perceptual Quality Assessment of Pan-Sharpened Images," Remote Sensing, vol. 11, no. 7, pp. 1-19, 2019.

[12] D.E. Moreno-Villamarin, H.D. Benitez-Restrepo, A.C Bovik, "Predicting the Quality of Fused Long Wave Infrared and Visible Light Images," Transactions on Image Processing, vol. 26, no.7, pp. 3479–3491, 2017.

[13] J. Guo, F. Yang, H. Tan, J. Wang and Z. Liu, "Image matching using structural similarity and geometric constraint approaches on remote sensing images," Journal of Applied Remote Sensing, vol. 10, no. 4, pp. 1-12, 2016.

[14] D. Liu, Y. Li, and S. Chen, "No-reference remote sensing image quality assessment based on the region of interest and structural

similarity," in Proceedings of the 2nd International Conference on Advances in Image Processing, Chengdu, China, pp. 64–67, 2018

[15] O. Ieremeiev, V. Lukin, K. Okarma, and K. Egiazarian, "Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images," Remote Sensing, vol. 12, no. 15, pp. 1-31, 2020.

[16] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Processing: Image Communication, vol. 30, pp. 57-77, 2015.

[17] O. Ieremeiev, V. Abramova, K. Okarma, and K. Egiazarian, "Improved robust linearized full-reference combined metric for remote sensing imaging," in Proceedings of the 6th Microwaves, Radar and Remote Sensing Symposium, Kharkiv, Ukraine, pp. 443-448, 2020.